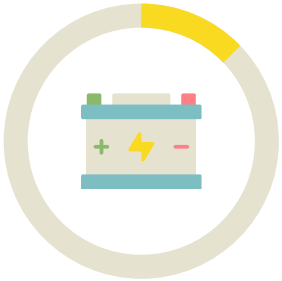
Problem Statement: Predict solar power potential in Namibia based on Global Horizontal Irradiance and meteorological data.

1. **Outcome variable:** Global Horizontal Irradiance in Namibia. (GHI_Avg)

2. Evaluate **16 variables** for significant prediction ability regarding meteorological data responsible for solar power potential in Namibia

3. Compare the performance of **7 models** for predicting solar irradiance

Data



Data Source:

Southern African Universities Radiometric Network

Solar Radiometric Data for the Public

SAURAN USAid Station in Namibia University of Science and Technology

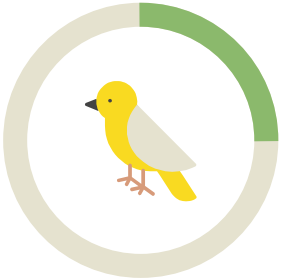
Data structure:

17 x 509,419

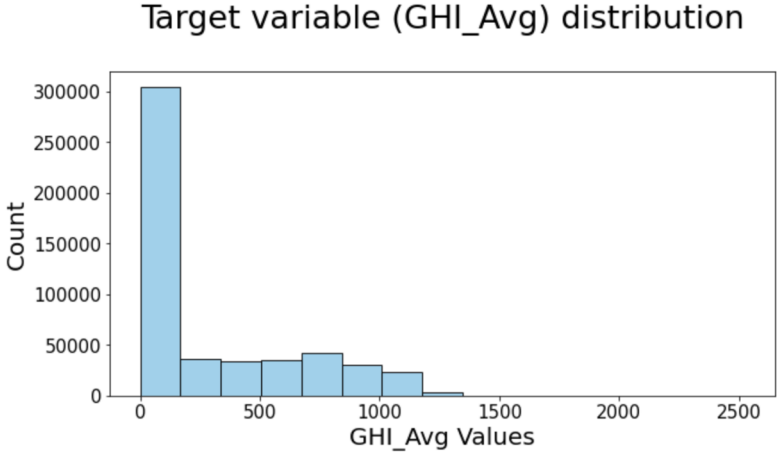
columns and rows pertaining to meteorological and physical conditions over a year.

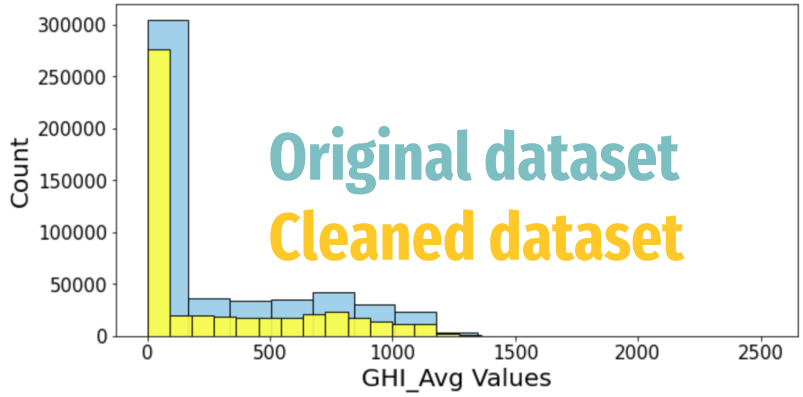
Data Processing/ Cleaning:

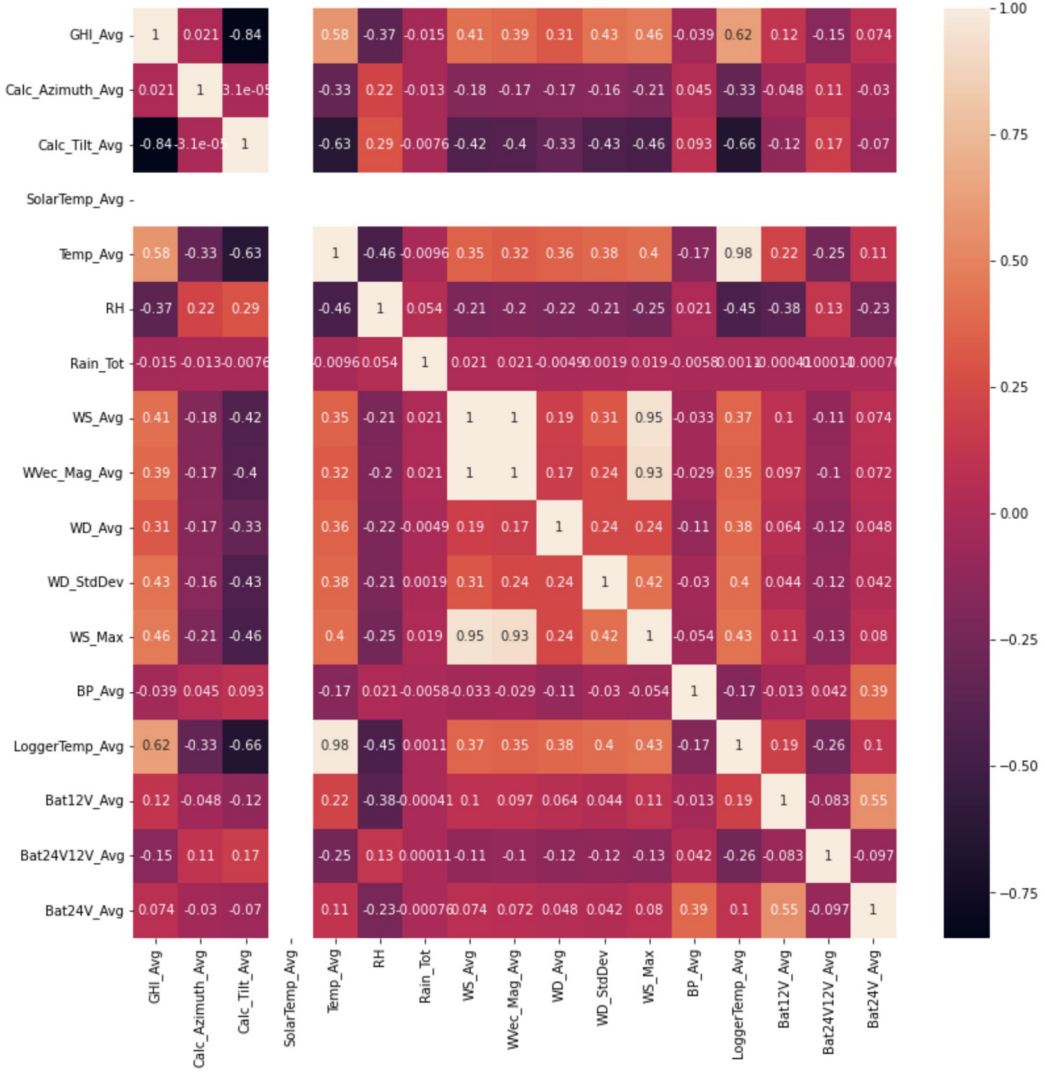
16 predictors with GHI (Global Horizontal Irradiance) as the outcome.



Exploring Data







Models applied

Linear Reg

Fitting the model
Analyzing intercepts and coefficients
Running predictions
Performing OLS Regression

KNN

Testing model with 2 neighbors
Using GridSearchCV to improve model
Evaluating and comparing models

Random Forest

Addressing potential multicollinearity and using correlation matrix
Using Variance Inflation Factor
Performing Random Forest Regression

Ridge

Introducing array of alpha with different values
Select optimal alpha by performing cross-validated ridge regression
Refit ridge regression with optimal value of alpha

Lasso

Fit the lasso model with 10000 maximum iterations
Perform 10-fold cross-validation to choose optimal alpha
Evaluate the model comparing ridge and lasso

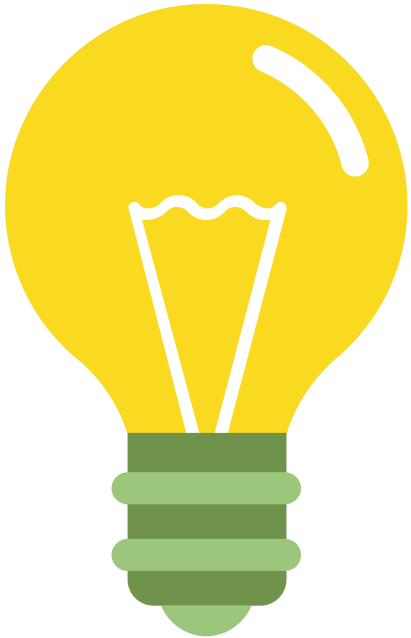
XGB

Perform training and testing on the dataset
Evaluate model and compare errors with previous models

GBR

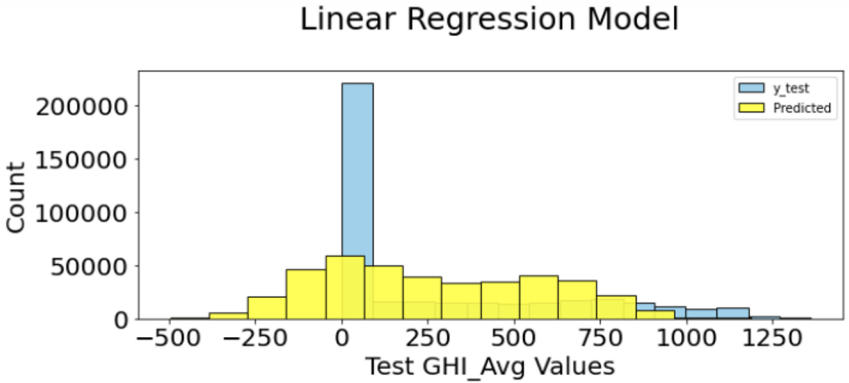
Instantiate gradient boosting regressor
Fit to training set
Predict on test set
Evaluate model

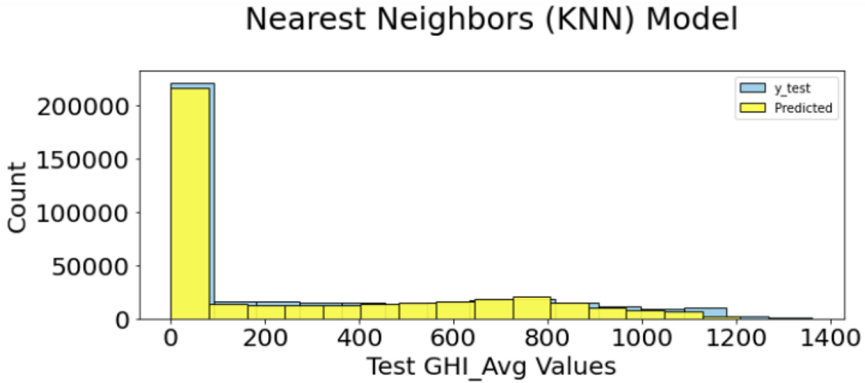
Model assessment

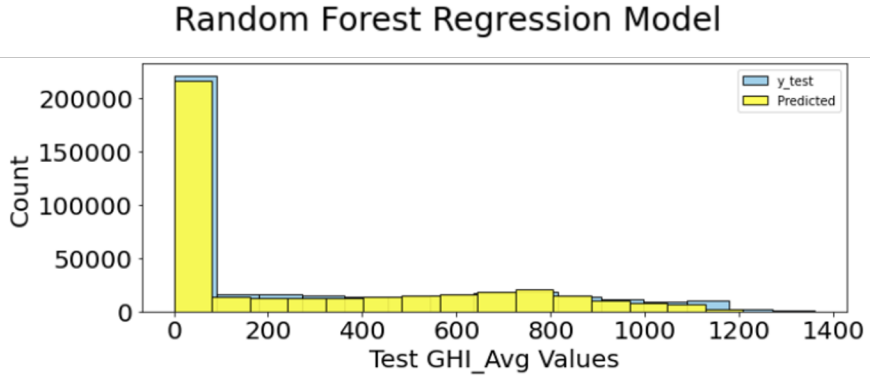


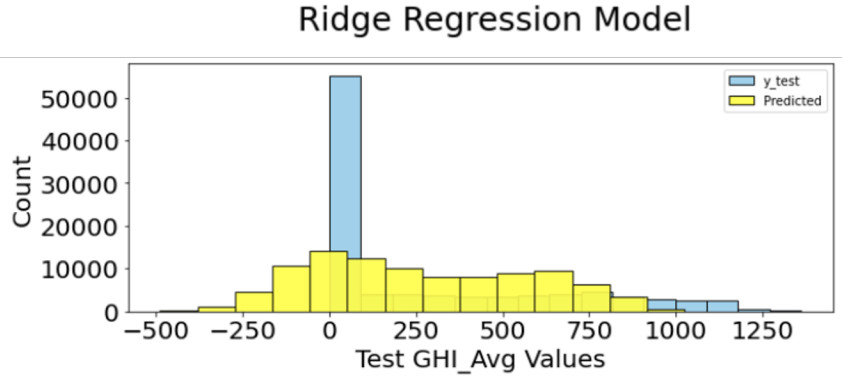
	r-squared	RMSE	MAE
Linear Regression Model	0.756	177.495	141.978
Nearest Neighbors (KNN) Model	0.928	96.285	37.66
Random Forest Regression Model	0.959	71.985	24.142
Ridge Regression Model	0.754	178.623	143.723
Lasso Regression Model	0.757	177.680	141.987
XGB Model	0.958	73.564	27.542
GBR Model	0.898	114.744	62.918

Model assessment



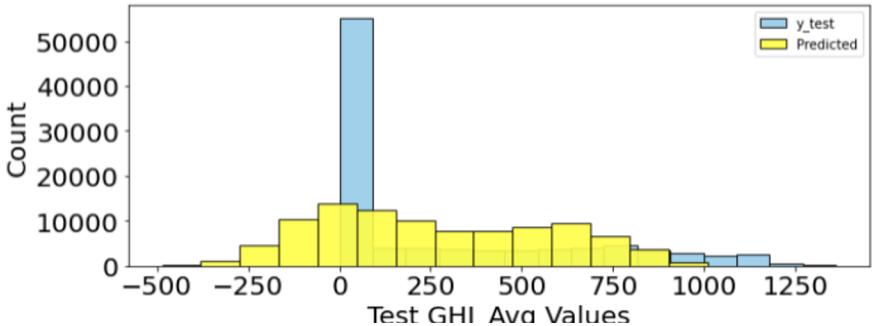




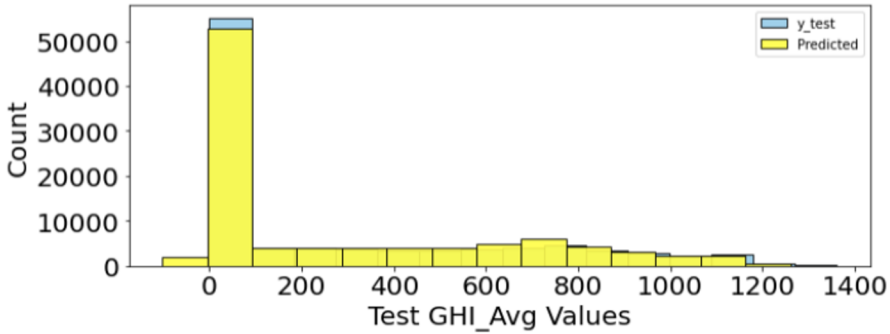


Model assessment

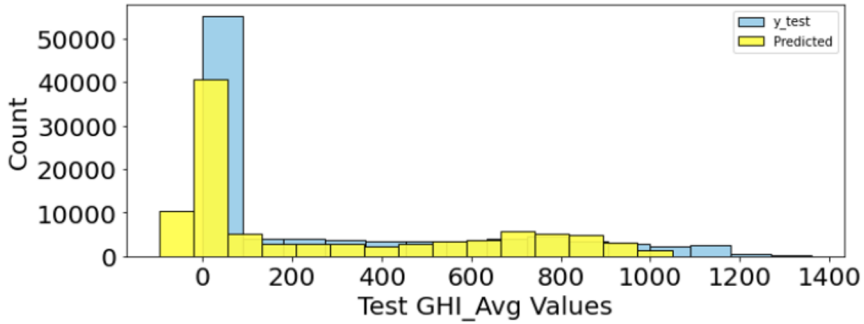
Lasso Regression Model



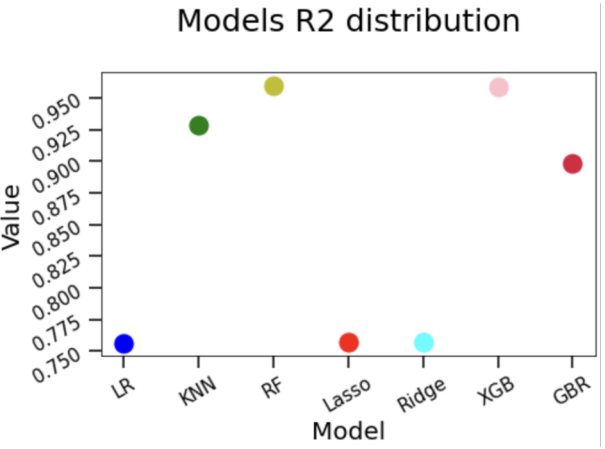
XGB Model

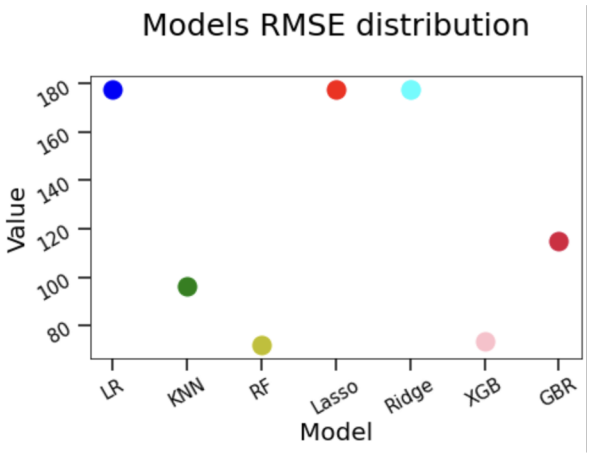


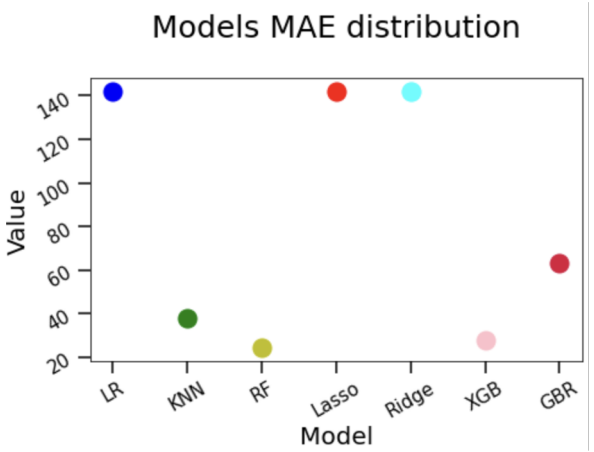
GBR Model



Model assessment







Random Forest and XGB

- *Random Forest (RF) and XGB are the two optimal models for our dataset.*
- *RF is slightly better than XGB (higher R2 and lower RMSE & MAE).*